

Adaptive Vocabulary Instruction

Samuel R. H. Joseph
*Laboratory for Interactive
Learning Technology,
University of Hawaii
srjoseph@hawaii.edu*

Andrew Smith Lewis
*Cerego Japan Inc., Tokyo
asl@cerego.co.jp*

Michael H. Joseph
*School of Psychology,
University of Leicester, UK
mhj1@le.ac.uk*

Abstract

In this paper we present some research into vocabulary instruction using a web-based program that allows subjects to interact with paired associates online. The system trains the subject on sets, or "lessons", of paired associates in a cue/response format. The sequence in which items are presented can vary according to a particular algorithm, and/or to the particular subject's ongoing performance. The system was used to study the acquisition of Japanese vocabulary in English speakers with no prior experience of Japanese. A two-way ANOVA indicated that there were significant effects on retention at test of lesson number, and of the algorithms controlling the order and selection of items at acquisition, and that there was no interaction between these factors. In this paper we consider the implications of these results for vocabulary instruction, and discuss how our system could be used to facilitate the study of other factors influencing learning and memory.

1. Introduction

There is extensive literature on the different aspects of vocabulary learning; with results showing that associating sentences with vocabulary items or requiring learners to perform generative tasks improves retention [4,8]. Other experiments have confirmed the widely known memory boosting effect of mnemonic strategies as well as indicating that a scripted pair-learning/testing format can provide additional benefits [7,9]. Conversely, some studies have indicated that visual repetition of vocabulary items correlates negatively with performance [6], and emphasize the positive effects of meta-cognitive strategies such as "Self Initiation" and "Selective Attention".

The broad range of studies on vocabulary learning over different nationalities, languages, ages, and so forth makes it difficult to find definitive support for one particular approach. In our current study we focus on an automated presentation paradigm in which we compare

the effects of presenting vocabulary items in different sequences, in some cases adapting that sequence flexibly to the learner in question. Any serious vocabulary learning system would naturally include support for example sentences, audio, encourage the use of mnemonics, etc., and we hope to study the various interactions of all these different factors in the future.

In section 2 we will look at the paired associate paradigm and its possible application to vocabulary learning. In the next section we examine the Atkinson model, particularly interesting for its use of a multiple state memory model; while in section 4 we look at some experiments that compare different ways of presenting vocabulary as paired associates. In section 5 we look at the results of these experiments. Finally, section 6 considers the implications of these results for online education systems, and how such systems would be structured to best support vocabulary learning.

2. Paired Associate Paradigm

The paired-associate learning task is a standard procedure for assessing human explicit memory. For example, randomly paired items such as words and letter strings are presented to subjects, who are then asked to recall one half of the pair from the corresponding other half. Rizzuto & Kahana [10] provide a summary of some different approaches to modeling of the paired associate learning task, as well as their own auto-associative neural network model. Rizzuto & Kahana focus on the extent to which different models can explain the differences between forward and backward recall of paired associates.

More directly relevant to our current topic of vocabulary learning is an approach presented by Atkinson and Crothers [2] that was based on a model that incorporated concepts of both short and long term memory. In their original study the predictions of different models of the day were compared with the results of a variety of different paired associate experiments, using tri-grams, Greek letters, digits and normal letters. Having demonstrated the explanatory

superiority of a three state model that distinguished between long-term and short-term memory as well as interference based forgetting, Atkinson [1] showed how the model could be applied to vocabulary learning.

3. Atkinson Model

The Atkinson Model takes a multiple state memory model that effectively distinguishes between long-term memory (LTM) and short-term memory (STM). LTM (state P) is considered permanent storage, but items in STM (state T) are temporary and may be forgotten, becoming unknown (state U). The assumption is that a learner will give a correct response when presented with any item that they have in either state P or T. Conversely, if that item is in state U they will give an incorrect response.

	P	T	U
P	1	0	0
T	x	1-x	0
U	y	z	1-y-z

Fig. 1: Presented Item Transition Matrix (see text for explanation)

	P	T	U
P	1	0	0
T	0	1-f	f
U	0	0	1

Fig. 2: Other Item Transition Matrix (see text for explanation)

The matrices in figures 1 and 2 show the probability of transition from one state to another with the left hand column being the state before presentation and the top row being the state after presentation. The presented item transition matrix in figure 1 is applied whenever an item is presented, e.g. if the presented item is currently in state U, then the likelihood of transferring to state P is y. The transition matrices are defined in terms of a number of parameters, x, y, z, and f which indicate how difficult to learn or easy to forget each item is. The second matrix

(fig. 2) is applied to those items that are not being presented, on each presentation that leads to an incorrect response. The implication is that interference from other items in the unknown state can cause an item to drop out of short-term memory.

Atkinson (1972) created a teaching system based on this model that always presented the item that was most likely to be transferred into state P. Given the sequences of correct and incorrect responses, the model would estimate the likelihood of each item being in a particular state. Given knowledge of the transition parameters, the system could then infer which item, if presented next, would most likely be transferred to state P. The assumption was that items in state P would remain in permanent store and thus be available for subsequent retention a week later. In experiments using English-German word pairs Atkinson's optimal strategy condition significantly outperformed subjects selecting their own study order (79% vs 58%). In addition there was a clear inverse relation between the performance during training, and subsequent test, e.g. the random condition subjects performed best during instruction, but much worse at test, while the optimal condition subjects performed badly during training but were the best during subsequent recall. If we think of the Atkinson optimal strategy as focusing on the most difficult items to ensure retention then this inverse relationship fits in with other results [11] indicating that things learnt under more demanding processing conditions are easier to remember.

Atkinson's procedure is complicated by the need to perform pilot studies on the same word pairs in order to calculate the values of x, y, z and f. A Chi-squared optimization was performed using a minimization algorithm to try and find the parameters that best fit the observed experimental data. A lack of detailed information about the procedures used in this modeling has made it difficult for us to replicate the algorithm precisely, which may account for our failure to fully replicate all of Atkinson's findings. We are currently working on a more complete modeling of the procedure with the collaboration of some of the original authors.

4. Experiments

In our own studies we attempted to replicate Atkinson's original findings, as well as comparing the Atkinson algorithm with another based on the expanded rehearsal series (ERS), explained below. The experimental work used a web-based program that allows subjects to study paired associates on-line. As described earlier the system was used to investigate the acquisition of Japanese vocabulary in English speakers with no prior experience of Japanese. The subjects were trained on sets, or "lessons", of vocabulary.

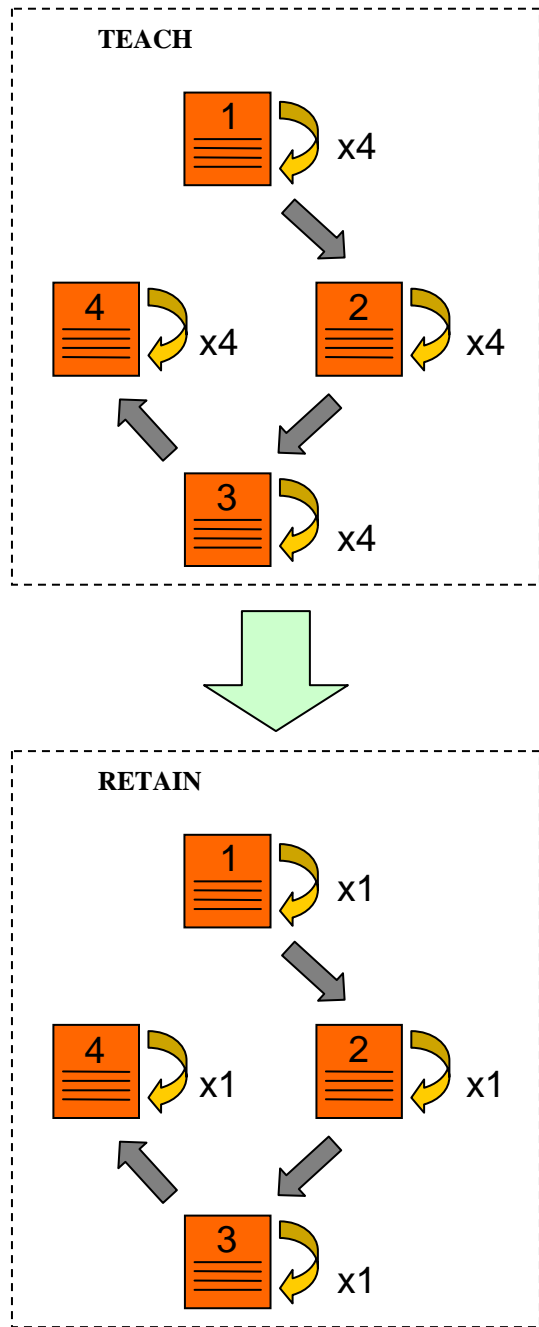


Fig. 3: TEACH and RETAIN lesson presentation.

The web-based system that we developed for our experiments can support a wide range of experimental conditions beyond those described here. Different algorithms can be plugged in to the system to control the sequence in which items are presented. After a cue has been displayed the subject is required to enter the appropriate response. In any one experiment, responses can be given in text-input (text entry matching allows for small typing errors) or multiple-choice formats, and

currently feedback is given after each choice. The experimenter can select from a range of lesson content as well as choose the duration for which lesson summaries are displayed to the subjects. The same web-based system supports subsequent test and each subject's individual activity is logged completely, and made available to the experimenter via a web interface. Only the number of internet-enabled computers available limits the number of subjects that can be run concurrently, since subjects can take part wherever they have access to such a computer.

The study described in this paper used text-entry and four 10-item lessons were presented in sequence, each repeated four times, with lesson summaries displayed for 30-second periods. The lessons were composed of common Japanese words and their English counterparts, divided into semantic categories such as greetings, transport and other common groupings. Retention was measured once, a week later.

As shown in figure 3, each lesson was presented four times over during the TEACH phase, with the presentation order within lessons controlled by one of three algorithms. The RETAIN phase which followed a week later consisted of a single loop through each lesson, with the individual items in random order.

In this study the presentation algorithms used were:

- 1) Random
- 2) Atkinson
- 3) Accordion (or Expanded Rehearsal Series, ERS)

4.1. Algorithm Details

1) Random - a randomized sequence of the items in the lesson under the restriction that each item would be shown at least once.

2) Atkinson - sequence varies according to a multi-state model of the probability of each item entering long term memory, based both on the intrinsic difficulty of the item, and on the individual's current performance

3) Accordion (ERS) - learned items are steadily further delayed before re-presentation, as long as they are not forgotten, in order to take advantage of the well-known spacing-effect [3,5]. Incorrect responses stop the re-presentation delay from increasing.

5. Results

In the present study, three groups, each of 10-12 undergraduate psychology students, in the School of Psychology, University of Leicester, UK, served as subjects. All subjects completed a demographic questionnaire including age, sex, and previous experience with foreign languages, including Japanese. Sex composition and mean age of the sample did not differ

significantly between groups. The mean age was 19.3 years, range 18-26.

5.1. Results at TEACH

A two-way ANOVA indicated that both "lesson", and "algorithm" had significant effects on TEACH scores; $F(3,116) = 5.48, p < .001$ and $F(2,116) = 4.69, p < .011$, respectively. However there was no significant interaction between these factors. TEACH scores were in the order $ATK > RAN = ACC$, while the absolute differences although noticeable were not very great (see fig. 4).

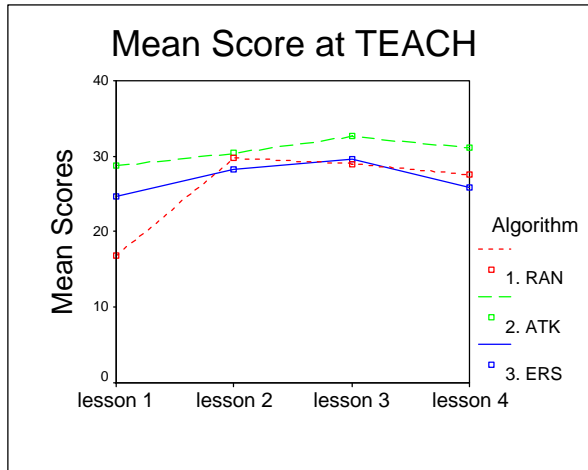


Fig. 4: The mean number of correct responses (y-axis) during the TEACH phase for each lesson (x-axis), for each different sequencing algorithm.

5.2. Results at RETAIN

A two-way ANOVA indicated that both "lesson" and "algorithm" also had significant effects on RETAIN scores; $F(3,116) = 6.98, p < .001$ and $F(2,116) = 8.59$, respectively, again with no significant interaction. However the RETAIN scores were inverted from the TEACH ones, i.e. in the order $RAN = ACC > ATK$ (see fig. 5)

5.3. Correlations

Over all subjects there was a positive correlation ($r = 0.475; p < 0.006$) between the TEACH and RETAIN scores, as might be expected. Within the different algorithm groups, positive correlations were also found, of 0.475 (RAN); 0.596 (ATK) and 0.861 (ACC). As these figures suggest, however, the mean or average TEACH and RETAIN scores for each group did show an inverse relationship between algorithms, as predicted by Atkinson.

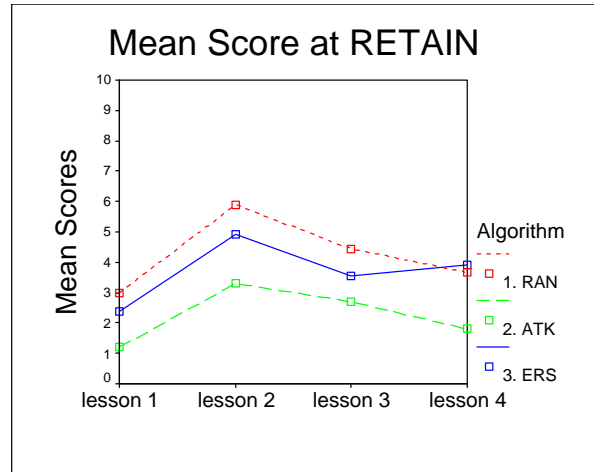


Fig. 5: The mean number of correct responses (y-axis) during the RETAIN phase for each lesson (x-axis), for each different sequencing algorithm.

The lesson effects are of unknown origin. Factors might include order effects (the design was not balanced for order) and inherent difficulty. It is not clear why the ATK algorithm performed relatively poorly, and the ERS algorithm was no better than RAN. One possibility is that the Atkinson acted in too extreme a fashion, resulting in the persistent presentation of a few items. Differences in the type of random algorithm used in the control condition may also have had an effect. Atkinson's original random condition had no constraints that required all items to be shown at least once. The implication is that our random condition outperformed our Atkinson algorithm due to the use of a better heuristic (ensure each item is presented at least once), and/or our inability to precisely match Atkinson's original.

Further analysis of the implementations of the non-random algorithms used, both of which have strong empirical support for their superiority, is indicated. Other directions that could be explored include looking at the permanence of state P assumption, which presumably must break down at some point. There is also the possibility of being able to adjust the state transition parameters in real time in response to subjects' performance. It seems likely that Atkinson's optimization process might benefit from the addition of longer term retention data, however it would be expedient to more fully replicate the original results before attempting to adjust the original model.

6. Discussion

The testing infrastructure developed to support our studies utilizes a thin client that performs minimal computation, and relies on a centralized server to handle all the operations necessary to the support of the different

sequencing algorithms. The advantage of this is that there is no need for the client to download special software, and in addition, proprietary algorithms can be tested without exposing their details unnecessarily. The implication of our sequencing studies so far is that parameters need to be carefully controlled in order to improve retention, but that certain heuristics can be employed to good effect.

There are many other factors to consider in order to create a vocabulary instruction system that will provide a satisfying user experience. In another recent study we have started investigating the effects of different entry methods on retention, with preliminary results indicating that text entry input at TRAIN, combined with multiple choice at TEST leads to increased performance at TEST. This again fits in with the suggestion that things learnt under more demanding processing conditions are more easily remembered [11].

Other factors to consider include, but are not limited to: longer term sequencing issues, contextualizing materials, and community collaboration. A successful implementation of the Atkinson framework could be extremely helpful for a community of learners. The community's ongoing interaction with different forms of content would allow the system to determine learning difficulty from one part of the community before using it to help another.

7. Acknowledgements

Many thanks to Ruth Edwards, Annie Giaro and Emily Host, third-year Psychology-with-Neuroscience students, at the University of Leicester, who carried out the experimental work during their degree projects.

8. References

[1] Atkinson, R. (1972) Optimizing the learning of a second language vocabulary. *Journal of Experimental Psychology*, 96, 124- 129.

[2] Atkinson, R. & Crothers, E. (1964) A comparison of paired-associate learning models having different acquisition and retention axioms. *Journal of Mathematical Psychology*, 1, 285-315.

[3] Dempster, F. (1988) The spacing effect: a case study in the failure to apply the results of psychological research. *American Psychologist*, 43 (8): 627-634.

[4] Grace, C. (1998) Retention of word meanings inferred from context and sentence-level translations: Implications for the design of beginning-level CALL software. *Modern Language Journal* 82 (4): 533-544.

[5] Greene, R (1989) Spacing effects in memory: evidence for a two-process account. *Journal of Exp. Psy.: Learning, Memory & Cognition*, 15 (3): 371-377.

[6] Gu, Y & Johnson R (1996) Vocabulary learning strategies and language learning outcomes. *Language Learning*, 46 (4): 643-679.

[7] Hansen, L., Umeda, Y. & McKinney, M. (2002) Savings in the relearning of second language vocabulary: The effects of time and proficiency. *Language Learning*, 52 (4): 653-678.

[8] Joe, A (1998) What effects do text-based tasks promoting generation have on incidental vocabulary acquisition? *Applied Linguistics*, 19 (3): 357-377.

[9] Jones, M., Levin M., Levin, J. & Beitzel, B. (2000) Can vocabulary-learning strategies and pair-learning formats be profitably combined? *Journal of Educational Psychology*. 92 (2): 256-262.

[10] Rizzuto, D. & Kahana, M. (2001) An autoassociative neural network model of paired-associate learning. *Neural Computation*, 13 (9): 2075-2092.

[11] Schneider, V., Healy, A. & Bourne, L. (2002) What is learned under difficult conditions is hard to forget: Contextual interference effects in foreign vocabulary acquisition, retention, and transfer. *Journal of Memory and Language*, 46 (2): 419-440.